

Revenue Management with Incomplete Demand Information

Victor F. Araman René Caldentey

Olayan School of Business, American University of Beirut, Beirut, Lebanon.
Stern School of Business, New York University, New York, NY 10012.

Abstract

Consider a seller who is endowed with a fixed number of units of a product that she can sell to a price-sensitive and stochastically arriving stream of consumers during a finite time horizon. The seller has incomplete demand information, that is, there are some characteristics of the demand process (*e.g.*, the arrival rate or the price elasticity) that she does not know with certainty. The seller's problem is to dynamically adjust the product's price to maximize the expected revenues she can collect if no replenishment is possible during the selling season.

Keywords: Revenue management, dynamic pricing, Bayesian learning, parametric and non-parametric learning, approximations, Poisson intensity control.

Consider a seller who is endowed with a fixed number of units of a product that she can sell to a price-sensitive and stochastically arriving stream of consumers during a given selling season. The seller's problem is to dynamically adjust the product's price to maximize the revenues she can collect if no replenishment is possible during the sales horizon. This setting is quite typical of many industries including among others, airlines (selling seats on a specific flight), hotels (booking rooms on a particular night) and retailers (selling seasonal merchandize). Often, these problems are labeled as *revenue management* problems since operational decisions are driven solely by revenues; inventory and capacity costs are sunk and incurred independently of changes in prices and/or number of units sold. The assumption of a fixed capacity is by no means critical if we consider that in most of these industries capacity is flexible only in the long run. Moreover, capacity decisions and price decisions take place on different time scales. Issues regarding the type of airplanes to schedule on a particular route, or the number of rooms to build in a given hotel, or the amount of seasonal merchandise to purchase from an overseas supplier are decided long before demand is realized and price policies are implemented.

In this paper, we discuss specifically the revenue management problem alluded to in the previous paragraph using a stylized mathematical formulation. We focus on the research that considers the case in which the seller has incomplete demand information. That is, there are some characteristics of the demand process (*e.g.*, the arrival rate or the price elasticity) that the seller does not know with certainty. Through this discussion we aim at summarizing some of the fundamental theoretical results of the existing literature. The model is simplistic but, we believe that the methods used and the insights drawn are quite representative of more general models and various other settings related to revenue management and, more broadly, to inventory management. Our exposition is

by no means exhaustive and we refer the reader to [11] and [29] for a comprehensive review of the literature on dynamic pricing and revenue management and to [14] for a detailed exposition on point processes and their optimal intensity control. We start reviewing the basic mathematical model under complete demand information.

1 Dynamic Pricing with Complete Demand Information

Consider a stream of potential customers arriving according to a time-homogeneous and price-independent Poisson process with rate Λ . We refer to Λ as the *market size*. Upon arrival a consumer buys the product with probability $\bar{F}(p)$, where p is the price of the product listed at this time. We can interpret $\bar{F}(p)$ as follows. We associate to each consumer a *reservation price* (that is, a maximum willingness to pay) which is distributed according to $F(\cdot)$ among the population of consumers (see [12] for more details), so that $\bar{F}(p) = 1 - F(p)$. We will assume that F admits a density f . The effective demand process or the sales process that results from the above description is a non-homogeneous Poisson process with rate $\lambda : \mathbb{R}_+ \rightarrow [0, \Lambda]$, a continuous, bounded and decreasing function such that at each time t , $\lambda_t = \lambda(p_t) = \Lambda \bar{F}(p_t)$, where p_t is the price charged at time t . We will refer to $\lambda(\cdot)$ as the *demand function*. Let \mathcal{A} be the set of *admissible* price processes $p = (p_t : t \geq 0)$, that is, processes for which the value of p_t depends exclusively on the history of sales and prices up to time t (but not on future unobserved events). We will denote by \mathcal{F}_t this history up to time t .

Let $\mathcal{N} = (\mathcal{N}_t : t \geq 0)$ be a standard (rate 1) Poisson process. For any $p \in \mathcal{A}$ and demand function λ , we define the cumulative demand process $N_\lambda^p(t) := \mathcal{N}(\int_0^t \lambda(p_s) ds)$ and its corresponding cumulative revenue

$$\mathcal{R}(p; x_0; \lambda) := \int_0^T p_t \mathbb{1}(N_\lambda^p(t) < x_0) dN_\lambda^p(t),$$

where T is the length of the sales horizon, x_0 is the seller's initial inventory and $\mathbb{1}(E)$ is the indicator function of the event E . The seller's dynamic pricing problem is given by the following (intensity control) problem

$$(P) \quad \sup_{p \in \mathcal{A}} \mathbb{E} [\mathcal{R}(p; x_0; \lambda)].$$

For $p \in \mathcal{A}$, Problem (P) can be rewritten as follows (see chapter II in [14] for details).

$$\sup_{p \in \mathcal{A}} \mathbb{E} \left[\int_0^T r(p_t) dt \right] \quad \text{subject to} \quad N_\lambda^p(T) \leq x_0 \quad (\text{a.s.}),$$

where $r(p) := p \lambda(p) = p \Lambda \bar{F}(p)$ is the instantaneous revenue rate. Let us define $p^* := \operatorname{argmax}\{r(p)\}$, $\lambda^* := \lambda(p^*)$ and $r^* := r(p^*)$. It is worth noticing that despite the fact that p^* maximizes the instantaneous revenue rate $r(p)$ choosing $p_t = p^*$ for all t is in general suboptimal. The reason is that this choice could deplete the available inventory too fast (before T) and the seller would be better off charging a higher price without necessarily sacrificing sales. On the other hand, if $x_0 = \infty$ then the seller can never deplete her inventory in finite time and $p_t = p^*$ for all $t \leq T$ is indeed an optimal policy.

A standard way to solve Problem (P) is by using dynamic programming. For this, let $J^*(x, t)$ be the seller's optimal revenue-to-go if the remaining sales horizon is $t > 0$ and the current available stock is x units. It follows that $J^*(x, t)$ satisfies the Hamilton-Jacobi-Bellman equation (see chapter VII in [14] for details)

$$\frac{\partial}{\partial t} J^*(x, t) = \max_{p \geq 0} \left[r(p) - \lambda(p) [J^*(x, t) - J^*(x-1, t)] \right],$$

with boundary conditions $J^*(x, 0) = J^*(0, t)$ for all $x, t \geq 0$. One can show that there exists a nondecreasing function ζ such that the optimal price satisfies $p^*(x, t) = \zeta(J^*(x, t) - J^*(x-1, t))$. Similarly, we define a nonnegative and decreasing function Ψ such that the right-hand side of the HJB is equal to $\Psi(J^*(x, t) - J^*(x-1, t))$. From the first-order optimality condition for $p^*(x, t)$ we get that the HJB leads to the following set of identities.

$$\frac{\partial}{\partial t} J^*(x, t) = \Psi(J^*(x, t) - J^*(x-1, t)) = \Lambda \frac{\bar{F}(p^*(x, t))}{h(p^*(x, t))}, \quad (1)$$

where $h(p) := f(p)/\bar{F}(p)$ is the hazard function of $F(p)$. It is worth noticing that the first equality defines an algorithm to solve the seller's optimization problem.

ALGORITHM:

- Step 0: Initialization. Set $J^*(0, t) = 0$ for all $t \in [0, T]$ and $n=1$.
- Step 1: Iteration. Given $J^*(n-1, t)$ for all $t \in [0, T]$ compute $J^*(n, t)$ as the solution of the first-order ordinary differential equation.

$$\frac{\partial}{\partial t} J^*(n, t) = \Psi(J^*(n, t) - J^*(n-1, t)) \quad \text{with border condition} \quad J^*(n, 0) = 0.$$

- Step 2: If $n = x$ then stop. Otherwise, set $n = n + 1$ and go to Step 1.

Closed-form solutions for $J^*(x, t)$ are not generally available. A notable exception is the case in which the reservation price is exponentially distributed, $\bar{F}(p) = \exp(-p/\theta)$. In this case, it can be shown that $J^*(x, t) = \theta \ln(\sum_{n=0}^x (\lambda^* t)^n / k!)$. For the general case, [19] obtain structural results for problem (P) using the HJB optimality condition.

Theorem 1 ([19]) *Suppose that $r(p)$ is a concave function. The revenue-to-go, $J^*(x, T)$, is strictly increasing and strictly concave in both T and x . Furthermore, there exists an optimal price $p^*(x, T) \geq p^*$ that is strictly decreasing in x and strictly increasing in T .*

It follows that inventory and time (the seller's primary resources) have diminishing marginal returns. The following result establishes the equivalence between the market size Λ and the sales horizon T and will be useful in our discussion on how to handle uncertainty on the market size.

Theorem 2 *Let $J_\Lambda^*(x, t)$ be the seller's revenue-to-go given a market size Λ . Then, $J_\Lambda^*(x, T) = J_1^*(x, \Lambda T)$ for all $x \geq 0$ and $T \geq 0$. It follows that, under the assumptions of Theorem 1, $J_\Lambda^*(x, T)$ is an increasing and concave function of Λ .*

PROOF: For any admissible pricing policy $p = (p_t : 0 \leq t \leq T)$, we define $\tilde{p} = (\tilde{p}_t : 0 \leq t \leq \Lambda T)$ such that $\tilde{p}_t = p_{t/\Lambda}$. The result follows by noticing that $\int_0^\tau \Lambda (1 - F(p_t)) dt = \int_0^{\Lambda\tau} (1 - F(\tilde{p}_t)) dt$ for all $\tau \in [0, T]$. Hence, the pricing policy p_t in $[0, \tau]$ generates (pathwise) the same revenue and depletes the same amount of inventory than the pricing policy \tilde{p}_t in $[0, \Lambda\tau]$. \square

Another key contribution in [19] is the observation that a fixed price policy can be asymptotically optimal as the scale of business (inventory and demand) increases. Denote by $p^D(x_0, T) := \max\{p^*, p^0\}$, where $\lambda(p^0) := x_0/T$. One can show that $p_t = p^D$ for all $t \leq T$, is an optimal policy for the deterministic version of Problem (P) (*i.e.*, one that replaces the stochastic increments dN_t^p by its deterministic counterpart $\lambda(p_t) dt$). We let $J^{\text{FP}}(x, t)$ be the seller's expected payoff if she selects the price $p^D(x, t)$ for the remaining sale horizon t .

Theorem 3 ([19]) *Suppose that $r(p)$ is a concave function. Then,*

$$1 - \frac{1}{2\sqrt{\min\{x, \lambda^* T\}}} \leq \frac{J^{\text{FP}}(x, T)}{J^*(x, T)} \leq 1.$$

It follows that when x and $\lambda^* T$ are both large a fixed price policy is sufficient to achieve an almost optimal revenue. This is a very useful result that limits the seller's needs for a possibly complex dynamic pricing policy.

An important limitation of the previous result is the assumption that the seller knows the specific demand function $p \mapsto \lambda(p)$. In practice, it is rarely the case that the seller can determine correctly this function. In some cases, some incomplete information is available based on historical data. In others cases, this demand function is completely unknown (*i.e.*, for new products or new markets). Furthermore, with incomplete demand information it is no longer true that structural properties of an optimal solution to Problem (P) will still hold. For example, with incomplete demand information there is no guarantee that a simple fixed price policy will produce asymptotically optimal results (*i.e.*, Theorem 3 is not guaranteed to hold). Indeed, as the following stylized example reveals an optimal fixed-price policy can perform very poorly if the seller does not know the true demand function.

EXAMPLE: Consider Problem (P) and suppose that at the beginning of the selling season the seller is uncertain about the true demand function $\lambda(p)$. She only knows that there is a 50% chance that $\lambda(p) = \lambda_i(p)$, where $\lambda_i(p) = \Lambda \mathbb{1}(p \leq p_i)$, $i = 1, 2$ for two fixed prices p_1 and p_2 such that $0 < p_1 < p_2 = 2p_1$. With full information, the seller knows the value of i and chooses an optimal pricing policy $p_t = p_i$ for all t with expected payoff $J_i^*(x, T) = p_i (\Lambda T - \mathbb{E}[(N(\Lambda T) - x_0)^+])$. Hence, it follows that with incomplete information the seller's payoff $V(x, t)$ is bounded by

$$V(x, t) \leq \frac{1}{2} J_1^*(x, T) + \frac{1}{2} J_2^*(x, T) = \frac{3}{4} p_2 (\Lambda T - \mathbb{E}[(N(\Lambda T) - x_0)^+]).$$

One can show that an optimal fixed-price policy (one that maximizes the seller's ex-ante expected revenue) is to set $p_t = p_2$ for all $t \leq T$. The corresponding expected payoff is given by

$$V^{\text{OFP}}(x_0, T) = \frac{p_2}{2} (\Lambda T - \mathbb{E}[(N(\Lambda T) - x_0)^+]).$$

Under this optimal fixed-price policy there is 50% chance that the seller would sell no unit during the entire selling season (*i.e.*, when $\lambda(p) = \lambda_1(p)$). Hence, the uncertainty on the true demand function and the lack of flexibility of a fixed-price policy to adjust prices over time make this fixed-price policy a risky and suboptimal strategy. To assess the degree of suboptimality, consider the following simple adaptive pricing policy designed to learn the true demand function from early sales. For a fixed $\alpha \in (0, 1)$, let $p_t = p_2$ for all $t \leq \alpha T$. If at least one sale occurs in $[0, \alpha T]$ then set $p_t = p_2$ for all $t \in [\alpha T, T]$, otherwise set $p_t = p_1$ for all $t \in [\alpha T, T]$. The expected payoff of this pricing policy is given by

$$V(x_0, T, \alpha) = \frac{(1 + e^{-\Lambda \alpha T})}{2} V^{\text{OFP}}(x_0, (1-\alpha)T) + \frac{(1 - e^{-\Lambda \alpha T})}{2} (p_2 + 2 \mathbb{E}_\tau[V^{\text{OFP}}(x_0 - 1, T - \tau)]),$$

where $\tau \in [0, \alpha T]$ is a random time with distribution $F_\tau(s) = (1 - e^{-\Lambda s}) / (1 - e^{-\Lambda \alpha T})$.

It follows that

$$\lim_{x_0 \rightarrow \infty} \frac{V(x_0, T, \alpha)}{V^{\text{OFP}}(x_0, T)} = 1 + \frac{(1 - \alpha)}{2} (1 - e^{-\Lambda \alpha T}) \xrightarrow{\tau \uparrow \infty} 1 + \frac{(1 - \alpha)}{2}.$$

According to Theorem 3, under full information a fixed-price policy is asymptotically optimal as x_0 and T grow large. However, in this case with incomplete information, the optimal fixed-price policy can deviate from an optimal strategy by as much as 50% (letting $\alpha \downarrow 0$) as x_0 and T grow large. At the same time, from the bound on $V(x_0, T)$ and the value of $V(x_0, T, \alpha)$ above, it follows that $V(x_0, T, \alpha) / V(x_0, T)$ converges to 1 as x_0 and T grow large and $\alpha \downarrow 0$. This result suggests that under incomplete demand information there could be a version of Theorem 3 stating that an almost everywhere fixed priced policy is asymptotically optimal, that is, a policy in which the seller needs to spend a small amount of time learning and then implementing a fixed price policy. We will see in Section 4.2 that such a Theorem does indeed exist under rather general conditions. \square

The previous example reveals the risks that a seller could take if she does not fully incorporate the effects of incomplete demand information when selecting optimal pricing strategies. The simple adaptive policy in this example also highlights the new informational role that pricing policies must play when there is uncertainty about the true demand function. With incomplete demand information the seller faces an *exploration-exploitation* trade-off, where the pricing decision the seller implements incorporates a learning component as part of the revenue maximization problem. In the following section, we summarize a set of theoretical frameworks that have been proposed to incorporate this incomplete demand information in the context of revenue management and dynamic pricing.

2 Dynamic Pricing with Incomplete Demand Information

In most (if not all) practical situations, the seller has only partial information about the true value of the demand function $\lambda(p)$. As a result, the dynamic pricing problem (P) needs to be modified to incorporate this additional *ambiguity* in the model formulation. Alternative solutions have been proposed to tackle this problem, which differ in two main characteristics: (*i*) the representation of the unknown function λ and (*ii*) the criteria that is used to resolve this model ambiguity.

Regarding the issue of how to model λ , two distinctive approaches have been studied: *parametric* and *non-parametric* models. As suggested by their names, the choice between these two alternatives largely depends on the capacity that the seller has to represent the demand function in terms of a finite number of parameters. To be concrete, suppose that the seller knows that $\lambda \in \mathcal{H}$, where \mathcal{H} is a given family of real-valued functions. If \mathcal{H} is the family of polynomials of degree K (for some fixed K) then $\lambda(p) = \beta_0 + \beta_1 p + \dots + \beta_K p^K$ and the seller's learning problem reduces to estimate the "best" set of parameters $\{\beta_k\}$. On the other hand, if $\mathcal{H} = \mathcal{C}_+[0, \infty)$, the set of nonnegative continuous functions in $[0, \infty)$, then selecting the "best" λ involves searching in this infinite-dimensional space; a problem that does not admit a parametric representation.

Independently of whether the seller uses a parametric or a nonparametric model to represent λ , the question of how to select an optimal pricing strategy needs to be addressed. This problem entails choosing a particular criteria to model how the seller's ambiguity about the true value of λ should impact her choice of prices.

The Bayesian approach is probably the most popular alternative used to model sequential demand learning for the case in which λ has a parametric description. Under this approach, model ambiguity on the demand function $\lambda(p)$ is captured by a set of unknown random parameters that characterize this function and for which a prior (at time 0) probability distribution is postulated. As time goes by and the evolution of the demand process is observed, this prior distribution is updated based on the new incoming information using Bayes rule. To be more specific, let θ be a random vector of parameters taking values in a set Θ and let $F(x)$ be the seller's prior probability distribution of θ . For each $\theta \in \Theta$ there is a corresponding demand function $\lambda_\theta \in \mathcal{H}$ and the seller's optimal dynamic pricing problem (P) takes now the following form

$$\sup_{p \in \mathcal{A}} \int_{\theta \in \Theta} \mathbb{E} [\mathcal{R}(p; x_0; \lambda_\theta)] dF(\theta).$$

From a modeling perspective, the Bayesian method relies heavily on the good judgment and expertise of the decision maker to select (i) the "right" functional form for the demand function and (ii) the "right" prior distribution for the unknown parameters. There are positive and negative sides to this approach. On the negative side, the parametric nature of the Bayesian model confines the demand learning within the boundaries specified by the proposed functional form and prior distribution and so any misspecification on these quantities will persist throughout the entire learning process. On the positive side, the Bayesian approach benefits from any prior knowledge that the decision maker may have about the demand process. This knowledge is particularly beneficial in those cases where the planning horizon is relatively short and there is limited time to learn through experimentation. Finally, Bayesian models can be mathematically tractable (with computationally efficient solution methods) for some specific family of distributions of the prior (those families of distributions for which their *conjugate* is tractable). In section 3.1, we review some representative papers that use this Bayesian method.

As we mentioned above, one valid criticism to the Bayesian approach is its strong dependence on the existence of a prior probability distribution for θ . Rather than modeling these unknown parameters as random variables, an alternative approach is to use a point estimate $\hat{\theta} \in \Theta$, where

the value of $\hat{\theta}$ is determined by optimizing a particular criteria that depends on θ and all available information. For example, at time $t = 0$, the *Maximum Likelihood estimation* of $\hat{\theta}$ solves

$$\hat{\theta} := \operatorname{argmax}_{\theta \in \Theta} \mathcal{L}(\theta, \mathcal{F}_0),$$

where $\mathcal{L}(\theta, \mathcal{F}_0)$ measures the likelihood of observing a history \mathcal{F}_0 for a given value of $\theta \in \Theta$. In other words, the seller's model ambiguity is resolved by mean of an optimization problem in terms of a likelihood objective function. Alternative estimation criteria have been proposed, essentially by replacing \mathcal{L} by other objective functions such as *Least Square estimation* (or more generally $\|\cdot\|_q$ -norm estimation), or *Minimum Entropy estimation*. As opposed to the Bayesian approach in which the exploration and exploitation stages are conducted simultaneously, in these models both steps are often decoupled. That is, given an estimate $\hat{\theta}$, the seller determines an optimal pricing strategy by maximizing expected revenue, $\mathbb{E}[\mathcal{R}(p; x_0; \lambda_{\hat{\theta}})]$. On the other hand, demand learning is achieved by periodically recomputing the value of $\hat{\theta}$ as new information is gathered, that is, replacing \mathcal{F}_0 by \mathcal{F}_t as time goes by. In section 3.2, we described a particular example that uses a Least Square estimation approach.

Another popular approach that has also been used extensively to handle model ambiguity is the class of *robust* formulations. The distinguishing feature of this approach is that it makes no probabilistic assumptions about which function $\lambda \in \mathcal{H}$ is more or less likely to be the true demand function. This *uncertainty set* \mathcal{H} (as it is usually called in this context) captures all the knowledge that sellers has about this demand. A robust pricing strategy is one that guarantees the best possible level of performance (*e.g.*, in a maximin, competitive ratio, or minimax regret criteria, among others) uniformly over all possible values of $\lambda \in \mathcal{H}$. For example, in the maximin version of problem (P), the seller chooses a dynamic pricing policy that solves for

$$\sup_{p \in \mathcal{A}} \inf_{\lambda \in \mathcal{H}} \mathbb{E}[\mathcal{R}(p; x_0; \lambda)];$$

that is, a pricing policy that maximizes the *worst-case* performance with respect to all demand functions in the uncertainty set \mathcal{H} . The popularity of this approach lies on the fact that it captures in a parsimonious way the seller's model ambiguity. In addition, and depending on the characteristics of \mathcal{H} (*e.g.*, polyhedral or ellipsoidal uncertainty sets), the maximin formulation can be solved efficiently. Among the disadvantages, the maximin criteria generally produces solutions that are too conservative, specially if \mathcal{H} is "big". For instance, suppose that \mathcal{H} contains a function λ_ϵ such that $\lambda_\epsilon(p) = 0$ for all $p \geq \epsilon$ for some $\epsilon > 0$ small. Then, an optimal maximin pricing policy satisfies $p_t^* < \epsilon$ for all t . It follows from this example that one needs to take special care in selecting the set \mathcal{H} under a maximin objective in order to avoid trivial or non-realistic solutions. To address this issue of conservatism, other robust criteria have been proposed. For example, under an absolute minimax regret approach the seller selects a pricing policy that minimizes the difference in performance with respect to the full-information case

$$\inf_{p \in \mathcal{A}} \sup_{\lambda \in \mathcal{H}} \left\{ \sup_{\tilde{p} \in \mathcal{A}} \mathbb{E}[\mathcal{R}(\tilde{p}; x_0; \lambda)] - \mathbb{E}[\mathcal{R}(p; x_0; \lambda)] \right\}.$$

Note that in this case, it is not generally true that an optimal pricing policy is conditioned by the presence of λ_ϵ in \mathcal{H} .

Another disadvantage of the robust approach –one that is particularly relevant in our present discussion– is that it does not explicitly incorporate the possibility of demand learning. In section 4.2, however, we review some recent results that show that in certain cases a simple policy that learns for a relatively small period of time and then myopically sets prices is asymptotically optimal in a robust sense.

In the following sections, we discuss some concrete models and examples within the class of parametric and nonparametric models for different demand learning criteria.

3 Parametric Models

Parametric models are based on some fundamental knowledge that the seller has about the true demand function. For example, the seller might know that $\lambda(p)$ belongs to a specific family of functions which is characterized by a finite set of parameters (e.g., the family of linear functions as in [7]). The seller parametric sequential learning problem is to identify the “best” value of those parameters using realized market information (prices and sales). As we mentioned above, there are different ways that one can use to identify the best set of parameters that we review in what follows.

3.1 Dynamic Pricing with Bayesian Learning

In the context of revenue management and dynamic pricing, there is a growing literature that uses Bayesian models to characterize optimal pricing policies when there is ambiguity about the true demand function. One particular type of uncertainty that has received significant attention is *market size uncertainty*. This model is well suited for those instances of the problem in which the seller knows relatively well consumers’ reservation price distribution $\bar{F}(p)$ but has limited information about the actual size of the population of potential buyers, Λ . Some representative papers that have studied this model are [3], [18] and [2] (see also [4] and [23]). In the context of Bayesian learning, [3] considers the special case in which Λ has a Gamma prior distribution and $\bar{F}(p) = \exp(-\alpha p)$ (*i.e.*, exponentially distributed reservation price). The advantage of using a Gamma distribution for Λ is that it is a conjugate distribution for the Poisson demand process which simplifies enormously the use of Bayes rule. In an infinite horizon setting, [18] generalizes the demand model in [3] to the family of finite mixture of Gamma distributions and propose a special heuristic (*decay balancing*) that shows a good numerical performance compared to other heuristics proposed in the literature. In the context of a retail operations, [2] considers an infinite horizon model for an arbitrary $\bar{F}(p)$ and where Λ has a finite support. A distinguishing feature in this model is that the seller has the option to optimally stop selling the product at any given time to switch to a different assortment. This is a particularly valuable option in the context of demand learning since the seller is initially uncertain about how profitable the product really is. For example, if the true value of Λ is small then the seller is better off removing the product.

In what follows we review the main results of the research discussed in the previous paragraph. The underlying assumption that we make in the remainder of this section is that Λ is the seller’s sole source of uncertainty. Having this in mind, we rewrite Problem (P). We denote by $G_0(\Lambda)$ the

seller's (prior) cumulative probability distribution of Λ at time 0 and by \mathbb{P}_0 (\mathbb{E}_0) the conditional probability measure (expectation operator) given G_0 . We define similarly G_t , \mathbb{P}_t and \mathbb{E}_t conditional on \mathcal{F}_t , the available information at time t . Using a slight abuse of notation, let $r(p) = p \bar{F}(p)$. The seller's problem is given by

$$(B) \quad \sup_{p \in \mathcal{A}} \mathbb{E}_0 \left[\int_0^T \Lambda r(p_t) dt \right] \quad \text{subject to} \quad \text{and} \quad \int_0^T dN_t^p \leq x_0 \quad (a.s.).$$

Similarly to Problem (P), we can solve Problem (B) using dynamic programming but with an enlarged state space (x, t, G_t) . The parametric description of G_t comes in handy at this point to ensure the tractability of the resulting DP algorithm.

The stochastic evolution of the new state variable G_t is derived from Bayes rule. For the sake of concreteness, let us suppose that Λ has a finite support $\{\Lambda_k\}_{k=1}^K$ (where $\{\Lambda_k\}$ is an increasing sequence) so that G_t is piecewise constant. Let $g_t(k) = G_t(\Lambda_k) - G_t(\Lambda_{k-1})$ with $\Lambda_0 = 0$. Then, Bayes rule implies that (see Proposition 3 and Section 6.1 in [2] for details)

$$g_t(k) = \mathbb{P}_0(\Lambda = \Lambda_k | \mathcal{F}_t) = \frac{g_0(k) \cdot (\Lambda_k I_t^p)^{N_t} \exp(-\Lambda_k I_t^p)}{\sum_j g_0(j) \cdot (\Lambda_j I_t^p)^{N_t} \exp(-\Lambda_j I_t^p)}.$$

It follows from its definition that $\{g_t, \mathcal{F}_t\}$ is a martingale. Application of Itô's lemma in the previous equation leads to the following SDE

$$dg_t(k) = \eta_k(g_t) (\lambda_t \bar{\Lambda}(g_t) dt - dN_t) \quad 1 \leq k \leq K,$$

where $\bar{\Lambda}(g) := \sum_k g(k) \Lambda_k$ and $\eta_k(g) := g(k) \left(\frac{\bar{\Lambda}(g) - \Lambda_k}{\bar{\Lambda}(g)} \right)$.

The resulting Hamilton-Jacobi-Bellman (HJB) optimality condition for Problem (B) is given by

$$\frac{\partial}{\partial t} V(x, t, g) = \max_p \left[\bar{\Lambda}(g) \bar{F}(p) [V(x-1, t, g-\eta(g)) - V(x, t, g) + \eta(g) \cdot \nabla_g V(x, t, g)] + \bar{\Lambda}(g) r(p) \right], \quad (2)$$

where $\nabla_g V(x, t, g)$ is the gradient of $V(x, t, g)$ with respect to g . The boundary conditions are $V(0, t, g) = V(x, 0, g) = 0$ and $V(x, t, e_k) = J_1^*(x, \Lambda_k t)$ where e_k is the distribution that gives probability one to the event $\{\Lambda = \Lambda_k\}$ and $J_1^*(x, t)$ is the full information value function for problem (P) when the market size is normalized to one. Note that the latter border condition follows from Theorem 2. Solving equation (2) is usually a very difficult task because of the singularities at the boundary points $g = e_k$ and the jumps in x and g that make the HJB a delayed difference-differential equation. Despite the fact that an analytical solution is not immediately available, this optimality condition provides enough information to derive some useful properties that we can use to approximate the value function and the corresponding pricing strategy.

Theorem 4 *The value function $V(x, t, g)$ satisfies the following properties:*

- i) *It is increasing in x and t . It is also convex in g .*
- ii) *It is bounded by: $J^*(x, \Lambda_1 t) \leq V(x, t, g) \leq \sum_k g(k) J^*(x, \Lambda_k t) \leq J^*(x, \bar{\Lambda}(g) t)$.*

iii) It satisfies $\lim_{x \rightarrow \infty} V(x, t, g) = \bar{\Lambda}(g) r(p^*) t = \lim_{x \rightarrow \infty} \sum_k g(k) J^*(x, \Lambda_k t)$. The optimal price satisfies $\lim_{x \rightarrow \infty} p^*(x, t) = p^*$.

iv) A fixed price policy is asymptotically optimal in the following sense. Let $V^{FP}(x, t, g)$ be the seller's expected revenue using an optimal fixed price for the entire sale horizon. Then,

$$1 - \frac{1}{2\sqrt{\min\{x, \bar{\Lambda}(g) \bar{F}(p^*) t\}}} \leq \frac{V^{FP}(x, t, g)}{V(x, t, g)} \leq 1.$$

PROOF: Part (i). The monotonicity of $V(x, t, g)$ on t and x follows trivially from its definition. The convexity on g follows from

$$V(x, t, g) = \sup_p \mathbb{E} \left[\sum_k g(k) \int_0^T p_t dN_t^p(k) \right], \quad \text{where} \quad N_t^p(k) := \mathcal{N} \left(\Lambda_k \int_0^t \bar{F}(p_s) ds \right).$$

Part (ii). The first inequality (lower bound) follows from implementing on the incomplete information case the optimal pricing policy for the case of full information with $\Lambda = \Lambda_1$. The second inequality follows from interchanging the sup and the summation in the equation above. The third inequality follows from Theorem 2. Part (iii) follows from noticing that for $x = \infty$, the optimal pricing strategy is $p_t = p^*$ which maximizes the instantaneous revenue rate (see the discussion in Section 1 after the definition of Problem (P)) independently of the value of Λ . Finally, Part (iv) follows from combining the upper bound on Part (ii) and the result in Theorem 3. \square

Note that $\mathbb{E}[J^*(x, \Lambda t)] = \sum_k g(k) J^*(x, \Lambda_k t)$ and so point (ii) asserts that an upper bound for the value function is given by the expected value (over the unknown market size Λ) of the full information value function. Point (iv) extends [19] result in Theorem 3 to the case with uncertain market size. This result suggests that the lack of asymptotic optimality of a fixed-price policy that we identified in Example 1 is mainly due to the seller's uncertainty about the buyers reservation price distribution rather than the actual market size.

3.1.1 Approximations and Heuristics

One approach that has been used to approximate the optimal pricing policy is to first get an approximation of the value function and then plug it in the HJB equation (2) to derive an approximating pricing policy. For any approximation $V^{\text{aprx}}(x, t, g)$ of the value function, the corresponding pricing policy is given by

$$p^{\text{aprx}}(x, t, g) = \zeta \left(V^{\text{aprx}}(x-1, t, g - \eta(g)) - V^{\text{aprx}}(x, t, g) + \eta(g) \cdot \nabla_g V^{\text{aprx}}(x, t, g) \right), \quad (3)$$

the function ζ was defined in Section 1. (It should be clear that the pair $(V^{\text{aprx}}, p^{\text{aprx}})$ does not solve equation (2).) This approach was used in [2] to derive an *asymptotic approximation*. This policy is based on the observation that points (ii) and (iii) in Theorem 4 suggest the use of the following approximation for the value function $\tilde{V}(x, t, g) = \sum_{k=1}^K g(k) J^*(x, \Lambda_k t)$ which is asymptotically optimal as x grows large. Let us denote by $\tilde{p}(x, t, g)$ the pricing policy that results from using $\tilde{V}(x, t, g)$ in (3).

An alternative approach to approximate an optimal pricing policy is to solve a modified version of the HJB. A popular example is the so-called *naïve policy*. This approximation assumes that at every state (x, t, g) the seller solves the full information HJB in equation (1) replacing the unknown

Λ by its expected value $\bar{\Lambda}(g)$ (see [3] for details). Another example is the *decay balancing heuristic* proposed by [18]. Although this policy was derived in an infinite horizon setting, the main idea can be directly extrapolated to the finite horizon case. The decay balance policy combines the asymptotic policy $\tilde{V}(x, t, g)$ and the HJB equation (2) (see also equation (1)) to propose a pricing policy $p^{\text{BD}}(x, t, g)$ solution to

$$\frac{\partial}{\partial t} V(x, t, g) = \bar{\Lambda}(g) \frac{\bar{F}(p^{\text{BD}}(x, t, g))}{h(p^{\text{BD}}(x, t, g))}.$$

Under the additional assumption that the reservation price distribution, $F(p)$, has increasing failure rate the solution $p^{\text{BD}}(x, t, g)$ is unique. A set of numerical experiments comparing the *asymptotic approximation*, the *naïve policy* and the *decay balancing heuristic* is reported in [18].

3.2 Dynamic Pricing for Linear Demand with Unknown Coefficients

The Bayesian method, as discussed in the previous section, relies heavily on the seller's prior distribution of the unknown parameters. In those cases, when no prior exists, the seller must rely on an alternative statistical estimation method, such as the least square estimator (LSE), the maximum likelihood estimator (MLE), the minimum entropy estimator (MEE) and others. Some representative examples are [7] (which model is discussed more in details below), [24] which considers also a linear price demand function and obtains approximate solutions using convex programming methods, and [16]. [16] considers a binomial demand which is relevant in the internet environment representing those that visited a website and ended up buying the item. Under, such demand model some parameters are unknown and a Monte Carlo simulation is used to quantify the trade-off between learning and revenue maximization. This simulation allows to measure the performance of few suggested heuristics among others, the one-step-look-ahead policy. Because of the numerical flavor, the method could apply to various estimation techniques in particular maximum likelihood and Bayes.

[7] considers a discrete variant of problem (P) in which the demand in every period n is given by

$$d_n = \beta^0 + \beta^1 p_n + \epsilon_n,$$

where the ϵ_n 's are iid $(0, \sigma^2)$ normal random variables. The parameters β_0, β_1 and σ are unknown. The estimates at period n of these parameters are computed using the LSE method

$$(\hat{\beta}_n^0, \hat{\beta}_n^1) = \arg \min_{\mathbf{r} \in \mathbb{R}^2} \sum_{s=1}^{n-1} (d_s - \mathbf{x}'_s \mathbf{r})^2 \quad \text{and} \quad \hat{\sigma}_n^2 = \sum_{\tau=1}^{n-1} \frac{(d_\tau - \hat{\beta}_n^0 - \hat{\beta}_n^1 p_\tau)^2}{n-3} \quad n = 4, \dots, T,$$

where, $\mathbf{x}'_s = [1, p_s]$. At the beginning of period n , the decision variable p_n is selected to maximize the total expected revenue-to-go. To compute this revenue-to-go, the seller needs to estimate the distribution of the demand in period $s \geq n$ (denoted by $\hat{d}_{s,n}$) conditional on the current estimates $(\hat{\beta}_n^0, \hat{\beta}_n^1, \hat{\sigma}_n)$. It follows that $\hat{d}_{s,n} = \hat{\beta}_{s,n}^0 + \hat{\beta}_{s,n}^1 p_s + \hat{\epsilon}_{s,n}$, where, $\hat{\epsilon}_{s,n}$ is a normally distributed random variable with mean 0 and standard deviation $\hat{\sigma}_{s,n}$; we use the notation $\hat{y}_{s,n}$ to denote the current period n estimate of a parameter y for a future period s , $s = n, \dots, T$. The most critical step in this approach (similar to the Bayesian case) is to obtain an iterative process that, given a state space, computes future estimates. It can be shown that the evolution of these estimates is given by

$$\hat{\beta}_{s+1,n}^i = \hat{\beta}_{s,n}^i + \hat{\epsilon}_{s,n} \cdot h_i(p_s, \sum_{\tau=1}^{s-1} p_\tau, \sum_{\tau=1}^{s-1} p_\tau^2),$$

for $i \in \{0, 1\}$, and

$$\hat{\sigma}_{s+1,n}^2 = \mathcal{H}\left(p_s, \hat{\beta}_{s,n}^0, \hat{\beta}_{s,n}^1, \sum_{\tau=1}^{s-1} p_\tau^2, \sum_{\tau=1}^{s-1} p_\tau, \sum_{\tau=1}^{s-1} p_\tau d_\tau, \sum_{\tau=1}^{s-1} d_\tau, \hat{\sigma}_{s,n}^2\right),$$

for some real functions h_i and \mathcal{H} (see [7] for details). These estimates are used in a DP formulation where the state space is given by the vector

$$\mathbf{X}_s = \left(x_s, \hat{\beta}_{s,n}^0, \hat{\beta}_{s,n}^1, \sum_{\tau=1}^{s-1} p_\tau^2, \sum_{\tau=1}^{s-1} p_\tau, \sum_{\tau=1}^{s-1} p_\tau d_\tau, \sum_{\tau=1}^{s-1} d_\tau, \hat{\sigma}_{s,n}^2\right),$$

where x_n is the inventory available in period n . At each period n , one solves the following DP with

$$J_T(c_T, \hat{\beta}_{T,n}^0, \hat{\beta}_{T,n}^1, \hat{\sigma}_{T,n}^2) = \max_{p_T} \mathbb{E}_{\epsilon_{T,n}} p_T \min \left\{ (\beta_{T,n}^0 + \beta_{T,n}^1 p_T + \epsilon_{T,n}), c_T \right\},$$

and for $s = n, \dots, T-1$,

$$J_s(\mathbf{X}_s) = \max_{p_s} \mathbb{E}_{\epsilon_n} [p_s \min \{ \hat{\beta}_{s,n}^0 + \hat{\beta}_{s,n}^1 p_s + \hat{\epsilon}_{s,n}, c_s \} + J_{s+1}(\mathbf{X}_{s+1})], \quad (4)$$

where the components of \mathbf{X}_{n+1} get updated based on the selected value of p_s , the corresponding demand estimate and the iterative processes described above (e.g. $c_{s+1} = c_s - \min \{ \hat{\beta}_{s,n}^0 + \hat{\beta}_{s,n}^1 p_s + \hat{\epsilon}_{s,n}, c_s \}$).

A few observations are in order. First, we note that for a fixed n , the sequence $(\hat{\beta}_{s,n}^i : s = n, \dots, T)$ made of the estimates at time n for future values of the parameter is a stochastic process and in particular a martingale. This is an expected fact as it measures the learning process. A valid question at this point is how to possibly quantify the learning. It will be hard to obtain a tractable formulation of the gap between the optimal values and those obtained through the estimation process, but a Monte Carlo simulation can help in this regard. Similarly to [16], one can solve for the optimal pricing policy under known parameters β_0 , β_1 and σ and then implement the previous pricing policy obtained through the learning process.

The main characteristic of this problem is the concurrent parameter estimation and the pricing optimization (active learning). If we decouple these two actions (move to passive learning), whereby the pricing decision made at time n is assumed not to affect the parameters estimates at time $n+1$, then the state space of the DP can be reduced to one dimension, that of the available capacity.

The setting of [7] is similar to the initial problem (P). The difference, beyond the fact that time is discrete, is the choice itself of a linear demand rate with a normally distributed error term independent of the price. From a practical standpoint, this model fits well with the popular approach of forecasting demand using linear regressions. Another advantage of this model compared to those discussed in the Section 3.1 is that the learning process includes the price elasticity (*i.e.*, an unknown reservation price distribution) as well as the market size. From a modeling perspective, the choice of a normally distributed demand could potentially lead to estimation errors if the probability of a negative demand is not negligible (*e.g.*, low moving items). One possible way to take care of this problem without losing tractability is to consider a log-normal where $d_n = a \exp(\beta^0 + \beta^1 p_n + \epsilon_n)$ so that by taking logarithm we recover a linear model (see [28] for a discussion of related demand model in the context of retail operations).

4 Non-Parametric Models: MaxiMin and MiniMax Methods

Non-parametric models are concerned with settings where the form of the demand function is not known. We distinguish two classes of non-parametric models. A first one is known as the maximin approach. Such criterion induces the optimizer, to maximize the worst case profit. It is a very conservative approach. This conservatism can be reduced by introducing a carefully chosen “budget” of uncertainty which limits the set of possible distributions or demand functions (e.g. [6] and [8]). Many papers have applied this method in operations management contexts (see for example [9]), while [22] applied it specifically to the setting discussed in this paper. The other class of non-parametric models is known as the minimax approach whereby a regret function is introduced that measures the performance of a policy relative to the performance of the optimal policy under a fully known demand function. This approach is less conservative than the maximin and has also been applied to various settings including recently to revenue management. We refer the reader to [5], [20], [26] and references therein. The first two papers consider a single resource allocation problem and the latter analyzes a network revenue management. In the last section of this paper, we discuss in some detail some of these approaches.

Learning is a dimension that is often missing in robust models. Indeed, most of the papers cited above, disregard learning and study an optimization problem that constrains the unknown demand function to some static, pre-specified, (non-parametrized) set. Having said that, there exists non-parametric models that do rely on exploiting demand realizations in order to learn the demand function and shape the solution of the optimization problem. Understandably, this approach happens to have more of an algorithmic flavor; see for instance [25], [17] and (more relevant to our setting) [21] and [10]. We discuss in more details the model of [10] that consider specifically problem (P), relies on learning while using a minimax formulation. The approach used in [22] will also be discussed below as an application of the maximin criteria applied to problem (P).

Consider again the general formulation (P) of the model. As mentioned in the introduction, the expected value is taken with respect to the probability measure defined on the Poisson process filtration. In both the Bayesian and the demand estimation approaches, the unknown parameters were replaced by their estimate and the probability measure was adjusted accordingly (e.g. in the Bayesian case, \mathbb{P} is replaced by \mathbb{P}_g while in the demand estimate case, \mathbb{E} is replaced by \mathbb{E}_ϵ). Consider now the case where the demand belongs to a non parametric class of functions. The fact that the demand intensity is unknown makes the seller’s maximization problem (P) ill-defined. In order to overcome this first challenge, a maximin or a minimax formulation has been proposed which transform the problem into a stochastic game. The seller still has to choose a pricing policy and in response, nature (or a clairvoyant adversarial agent) picks the “worse” possible demand function given the seller’s choice. The (real) demand function that is unknown is replaced by the solution of an optimization problem. In other words, the seller selects the pricing policy to guaranty somehow the best worst-case performance.

In mathematical terms, the seller’s maximin formulation of problem (P) is given by

$$(M) \quad \sup_{p \in \mathcal{A}} \inf_{\lambda \in \mathcal{H}} \mathbb{E} \left[\int_0^T p_t \lambda(p_t) dt \right] \quad \text{subject to} \quad N_\lambda^p(T) \leq x_0 \quad (\text{a.s.}),$$

where \mathcal{H} is the (possibly infinite dimensional) set of functions that contains the true demand function.

4.1 MaxiMin and the Relative Entropy Formulation

A maximin approach as presented in Section 2, induces the seller to select the price that will maximize the worst case expected total revenues: a worst case with respect to the demand functions and the pricing policy adopted. Reducing the conservatism of this approach entails reducing the uncertainty set or the set of possible demand functions. We follow here the approach in [22], to suggest a robust formulation of Problem (P).

In the setting of problem (P), there is a one-to-one relationship between the demand function $\lambda(\cdot)$ of the Poisson process and the probability measure \mathbb{P} governing the system. Hence, the set of demand functions can be replaced by a set of probability measures and it is in the interest of the seller to pick carefully this set as it will dictate how conservative the solution will be.

Often it is the case that the seller has collected information (through historical data) on the “real” demand function she is or will be facing. This information is represented as a demand function, $\beta(\cdot)$, that is not necessarily the real (and still unknown) demand function $\lambda(\cdot)$. However, with some confidence one can assume that both demand rates are not too “far” apart. We denote by \mathbb{P} (respectively \mathbb{Q}) the probability measure induced by a demand rate $\lambda(\cdot)$ (respectively $\beta(\cdot)$). The set of probability measures \mathbb{P} close to \mathbb{Q} , from which a clairvoyant will draw the worst possible measure, is defined as follows:

- \mathbb{P} is absolutely continuous with respect to \mathbb{Q} , denoted by $\mathbb{P} \ll \mathbb{Q}$.
- \mathbb{P} satisfies $\mathcal{E}(\mathbb{P} | \mathbb{Q}) \leq \gamma$, where $\mathcal{E}(\mathbb{P} | \mathbb{Q}) := \mathbb{E}_{\mathbb{P}} \ln \eta(T)$ and $\eta(T)$ is the Radon-Nikodym derivative of \mathbb{P} with respect to \mathbb{Q} (see [14]). The function \mathcal{E} is known as the *relative entropy* of \mathbb{P} with respect to \mathbb{Q} and γ is some positive constant (sometimes refer to as the *budget of uncertainty*) that captures the seller’s confidence level about how close the true probability measure is to the nominal measure \mathbb{Q} (in a relative entropy sense).

These two conditions define an *uncertainty set* of probability measures. Note that for $\gamma \equiv 0$, this set reduces to $\{\mathbb{P} \equiv \mathbb{Q}\}$.

We consider the following transformation of problem (P) and a variant of problem (M) under a maximin criterion

$$J(x, T) = \sup_p \inf_{\mathbb{P} \ll \mathbb{Q}} \mathbb{E}_{\mathbb{P}} \int_0^T p(t) dN(t) \quad \text{subject to} \quad \mathcal{E}(\mathbb{P} | \mathbb{Q}) \leq \gamma,$$

Using Girsanov’s Theorem applied to point processes (see [14]), one can prove that there is a stronger correspondence between probability measures in the uncertainty set defined above and their intensities. In particular, the Poisson process with rate $(\lambda(p_t) : t \geq 0)$ under the (unknown) probability measure \mathbb{P} is a Poisson process under the nominal probability measure \mathbb{Q} but with an intensity $(\beta(p_t) \kappa(t) : t \geq 0)$ where $\kappa(\cdot)$ is some (unknown) non-negative process. Hence, instead of picking \mathbb{P} , it is enough for nature to pick κ under \mathbb{Q} . With this alternative formulation, one can solve the seller’s problem using dynamic programming. The corresponding HJB equation is given by

$$\frac{\partial}{\partial t} J(x, t) = \max_p \min_{\kappa} \left\{ \lambda(p) [p\kappa + \theta(\kappa \log \kappa + 1 - \kappa)] + \lambda(p)\kappa [J(x-1, t) - J(x, t)] \right\},$$

with border conditions $J(x, 0) = J(0, t) = 0$ and where the constant θ is fully characterized given γ (see [22] for details). One interesting feature of this equation is the interchangeability of min

and max under strong concavity of the nominal revenue rate. Moreover, as in [19] a closed-form solution can be obtained for an exponential demand function (see [22] for details).

4.2 MiniMax and the Regret Function

The recent work of [10] studies a minimax variant of problem (M). In particular, it considers a rather general non-parametric uncertainty set \mathcal{H} , which is the set of uniformly bounded and Lipschitz continuous functions with bounded support. A relative regret function, \mathcal{M} , is defined as one minus the ratio between the value function, $J^p(x, t; \lambda)$ (generated by any pricing policy p) and the value function in the deterministic case, $J^D(x, t|\lambda)$ (conditioned on knowing the true demand function λ),

$$\mathcal{M}^p(x, t; \lambda) := 1 - \frac{J^p(x, t; \lambda)}{J^D(x, t|\lambda)}.$$

This regret function is inspired by two results in [19] (see Theorem 3) that establish that (i) the value function under a deterministic demand upper bounds the value function under a Poisson demand and (ii) a fixed price policy, solution of the deterministic problem, is asymptotically optimal. Hence, the regret function is non-negative and bounded by 1 (independently of λ), and measures the performance of a pricing policy p for a specific demand function λ . The smaller the regret is, the better the pricing policy is, taking the value zero only when the pricing policy is able to achieve the value of the deterministic setting. At this point, it remains to transform the initial problem formulation into a stochastic game, where the seller chooses the “optimal” pricing policy taking into consideration that nature will then pick the worst possible demand curve (i.e. by maximizing the regret). The minimax *relative* regret formulation is written as follows

$$\inf_{p \in \mathcal{P}} \sup_{\lambda \in \mathcal{H}} \mathcal{M}^p(x, t; \lambda).$$

For any pricing policy p the inner maximization problem is (at least in theory) well defined. Solving the previous problem is in general not possible. One approach would be to construct a pricing policy that performs well. As we recalled in Theorem 3, when the demand function is known, say $\tilde{\lambda}$, a fixed price policy $p_t = p^D(\tilde{\lambda})$ is asymptotically optimal in the context of [19]. In the case where the demand curve is unknown, [10] proposes a policy that aims at estimating first the demand function λ and then fix the price to be equal to $p^D(\lambda)$. Specifically, they explore first how demand reacts to a set of selected prices (passive learning phase) and then exploit this learning to get approximations of both the run-out price, p^0 and the maximizer of the demand function, p^* . (See the discussion that precedes Theorem 3 for the definitions of p^0 and p^*). In the second phase, they adopt a fixed price policy which is the minimum of the two approximated prices. The longer the learning phase is, the more accurate the price approximations are; but the higher the opportunity cost is (due to the non-optimal pricing during the learning phase). Algorithm 1 is sensitive to the length of the learning period τ and to a granularity parameter κ and follows the following steps. 1. In the *initialization* step κ prices are picked equidistantly from the interval of possible prices $[\underline{p}, \bar{p}]$. The interval $[0, \tau]$ is divided in κ equal time intervals, $\Delta = \tau/\kappa$. 2. In the *learning/experimentation* step, each of the κ pre-selected prices is applied on one of the κ intervals of time of length Δ . Demand gets realized accordingly and at the end of this phase, the total demand realized in each of these intervals is computed; each demand value is divided by Δ thus representing an approximation of the demand rate function $\hat{d}(p_i)$ at the price selected for that interval. The approximated demand rate is in turns multiplied by the price and the result, $\hat{d}(p_i) \cdot p_i$, is an approximation of the revenue

rate at that price. At the end of these two steps, κ points approximating the demand and the revenue rate functions have been gathered. 3. The *optimization* step identifies two specific prices

$$\hat{p}^u = \arg \max_{1 \leq i \leq \kappa} \{\hat{d}(p_i) \cdot p_i\}, \quad \text{and} \quad \hat{p}^c = \arg \min_{1 \leq i \leq \kappa} |\hat{d}(p_i) - x/T|.$$

4. In the last *pricing* step, the price $\hat{p} = \hat{p}^u \wedge \hat{p}^c$ is applied on the interval $(\tau, T]$. We denote by $p(\tau, \kappa)$ the output of Algorithm 1. In terms of performance, this policy is shown to be asymptotically optimal in the following sense.

Proposition 1 *Set $\tau_x = x^{-1/4}$, $\kappa_x = x^{1/4}$ and let $p_x := p(\tau_x, \kappa_x)$ be given by Algorithm 1. Then the sequence $\{p_x\}$ is asymptotically optimal, and for all $x \geq 1$*

$$\frac{C'}{x^{1/2}} \leq \sup_{\lambda \in \mathcal{H}} \mathcal{M}_x^p(x; \lambda) \leq \frac{C(\log x)^{1/2}}{x^{1/4}},$$

for some positive constant C and C' .

The lower bound (in fact obtained under slightly stronger assumptions) measures the decreasing gap between Algorithm 1 and the optimal solution.

Clearly, this approach can also be applied to the case where the demand function belongs to a parameterized set $(\lambda(\cdot; \eta) \in \mathcal{H}_\eta)$, as it was the case in both the Bayesian and the demand estimate methods. It is easily shown that the pricing policy generated by Algorithm 1 remains asymptotically optimal. The example of Section 1 is a good illustration of such a setting and how effective is a pricing policy driven by a short learning period followed by a fixed price policy. Moreover and as expected, Algorithm 1's performance improves in the parameterized case. The corresponding upper bound of Proposition 1 is tighter and the number of pricing test points during the learning phase is set equal to the number of unknown parameters (i.e. bounded and generally small). Such an interesting result requires that the parameterized demand function satisfies additional regularity assumptions. It needs to be "identifiable" based on a set of observations, which in particular means that for any vector $d = (d_1, \dots, d_k)$, the system of equations $\{\lambda(p_i; \eta) = d_i, i = 1, \dots, k\}$ has a unique solution in η .

Non parametric approaches relying on robust controls, as we saw above, do not follow so much a general framework or methodology as much as their parametric counterparts do (e.g. Bayesian or statistical estimation methods). Non parametric approaches are stemmed essentially from the specific model under study. In order to have a broader sense of the different techniques that could be used, we discuss in the next section another revenue management setting that has recently been analyzed in the context of robust controls. Again, the main drawback of these approaches is the lack of learning. They do induce robust decisions, however the realizations of demand are not being incorporated in any way to reduce the uncertainty set.

4.3 Robust Single-Leg Revenue Management Model

The so-called single-leg model is to revenue management what the newsboy model is to inventory management. (A broad literature is available on this subject and is very well summarized in [29].) The setting in this single-leg revenue management model is almost identical to the dynamic pricing model that we have discussed so far with one notable difference; items can be sold at different prices at the same time. The canonical example corresponds to seats in an airplane that, depending on the

different restrictions that the airline imposes on the products (*e.g.*, refundable vs. non-refundable tickets), are sold at different fares. These fare classes are typically announced at time 0 and remain fixed throughout the entire selling season. In this setting, the seller, rather than controlling a single price over time, selects dynamically which fare classes to *open* and which ones to *close* as a function of the available inventory and remaining sale horizon. Non parametric methods have been used recently in the literature to formulate and solve this “cousin” version of Problem (P).

As before, let x_0 be the seller’s initial inventory and let $f_1 > f_2 > \dots > f_m$ be the fares associated to the m different products offered by the seller. We assume that every product consumes one unit of inventory. (In the general setting of *network revenue management*, the seller is endowed with a collections of different resources and different products consume a combination of resources in different quantities.) The seller’s objective is to select an admission control policy π that specifies which fare classes to accept (open) and which one to reject (close) based on the state of the system at every point in time. A renowned strategy is the *standard nesting* policy which works as follows. The seller selects *booking limits* $(\theta_1, \theta_2, \dots, \theta_m)$ associated to the m fare classes, such that $\theta_1 \geq \theta_2 \geq \dots \geq \theta_m$. A class- j request is accepted (*i.e.*, to whom one unit of resource is sold at f_j) only if the number of customers of classes j to m (*i.e.*, corresponding to fare f_j or lower) already accepted is less than or equal to θ_j .

[5] analyzes this problem in an environment where demand information is incomplete using critical ratio ideas, which are popular in the context of competitive analysis for online algorithms (see [1] for a survey on the subject). We denote by $R^\pi(I)$ the revenues generated by applying some policy π to a stream of arrivals, I . We call this an “online algorithm”. We let $R^*(I)$ be the revenues obtained by applying the best possible policy for that particular stream of arrivals I . This is the “offline algorithm” revenues that can be achieved if the seller selects the policy after she sees the entire stream of arrivals. An online algorithm can be deterministic (*i.e.*, for a specific stream of demand the output is deterministic) or randomized (*i.e.*, for a specific stream of arrivals, the output depends on the outcome of some random variable). We refer the reader to [20] for a brief discussion of randomized policies.

We let Ω_π be the set of all possible demand streams I that a policy π can faced and define the competitive ratio as

$$\Upsilon(\pi) := \inf_{I \in \Omega_\pi} \frac{\mathbb{E}R^\pi(I)}{R^*(I)}.$$

We denote by c^* the supremum of Υ on all possible (deterministic or randomized) policies π . It is not hard to see that by maximizing the competitive ratio we are minimizing the worst case relative regret $\sup[1 - \mathbb{E}R^\pi(I)/R^*(I)]$. We should also note that computing c^* requires a pathwise deterministic optimization as the expected value is only with respect to the possibly randomized algorithm π . This relative regret formulation is different from the one defined in [10]. In the latter the denominator of the regret function was the value function under a deterministic demand. In a competitive ratio setting, the denominator is a function of the specific stream of arrivals. These two denominators are asymptotically the same in the case of Problem (P) when the demand function is known.

The competitive ratio of a particular policy π , measures its performance which is to be compared to c^* . In particular, a policy that achieves c^* is optimal. [5] uses this method to evaluate c^* for a

multi-fare single leg problem and shows that

$$c^* = \left(m - \sum_{j=2}^m \frac{f_j}{f_j - 1}\right)^{-1}.$$

It is also shown that standard nesting policies achieve c^* (and hence are optimal under a competitive ratio criterion) when the (nested) booking limits, θ_j 's are given by

$$\theta_j = x_0 - \frac{x_0}{c^*} \left(j - \sum_{i=1}^j \frac{f_{i+1}}{f_i}\right).$$

The approach adopted to prove such result is as follows. One first recognizes the optimal policy, (e.g. the standard nesting) and calculates its optimal competitive ratio (e.g. find the best possible booking limits). It remains then to show that this competitive ratio is the best achievable among all other (deterministic or randomized) policies. The first step is proven by dissecting all possible arrival streams with respect to the candidate policy. The most critical step is the second one that often relies on a powerful result, the Neuman/Yao principle, which states that

$$c^* \geq \sup_{\pi_d} \mathbb{E}_Q \frac{R^{\pi_d}(I_Q)}{R^*(I_Q)},$$

where I_Q is a random stream of arrivals chosen according to any probability distribution Q , and π_d is any deterministic online algorithm. Basically, it shifts the search for the optimum from a pathwise search under both randomized and deterministic algorithms to a probability distribution on the stream of arrivals under deterministic algorithms. (See [27] that provides a framework to prove bounds using the aforementioned principle). [20] builds on the competitive ratio approach of [5] to solve, among other things, for the case where the demand for each fare has known bounds.

Instead of using a competitive ratio criterion (equivalently a minimax relative regret) one can use other criteria such as the absolute minimax regret or the maximin criteria as defined in the previous section. As a matter of fact, for the multi-fare single leg problem [20] proves under an absolute minimax regret criterion that nested policies are again optimal. They propose a unified framework for the competitive ratio and the minimax absolute regret and prove that the absolute regret booking limits are more aggressive than those obtained through the competitive ratio criterion. Finally, [26] applies both a maximin and the absolute minimax criteria to a network revenue management problem where demand belongs to a polyhedral uncertainty set (generalizing the single-leg model in [20]). In this more complex setting they restrict their attention to a set of control policies that include the nested booking limits and develop robust formulations.

5 Summary

In this paper, we have reviewed a representative model of revenue management and discussed different methods to deal with incomplete demand information. Each method depends on the specific context and the level of information available. Parametric models open the way for active learning approaches in which demand realizations are used to update estimates of the unknown parameters over time. These estimates are generally embedded within a dynamic programming formulation. The solution of the DP characterizes the pricing policy to adopt. Non-parametric

models are less tractable. A first method that we discussed is based on the idea of replacing the demand uncertainty by a stochastic game using a maximin approach. In order to reduce the conservatism of this formulation, robust optimization is used limiting the feasible set. Eventually, the pricing policy is obtained by solving a complex ODE. An alternative non-parametric model uses passive learning during a specified interval of time followed by a fixed price policy for the remaining of the horizon. The pricing solution is simple to implement and asymptotically optimal. We finally discussed the competitive ratio criteria applied to the setting of a capacity allocation revenue management problem and compared that to the other robust criteria available in the literature.

References

- [1] Albers, S., Online algorithms: a survey. *Math. Programm.*, 2003, 97(1-2): 3-26.
- [2] Araman, V., Caldentey, R. Dynamic Pricing For Non-Perishable Products With Demand Learning. *Operations Research*, 2009, 57(5): 1169-1188.
- [3] Aviv, Y., Pazgal, A. Pricing of Short Life-Cycle Products through Active Learning. Working Paper, Olin School of Business, Washington University, St. Louis, 2002.
- [4] Aviv, Y., Pazgal, A. A Partially Observed Markov Decision Process for Dynamic Pricing. *Management Sci.*, 2005, 51(9):1400-1416.
- [5] Ball, M., Queyranne, M. Toward robust revenue management: Competitive analysis of online booking. *Operations Research*, 2009, 57(4): 950-963.
- [6] Ben-Tal, A., Nemirovski, A. Robust convex optimization. *Math. Oper. Res.*, 2003, 23:769-805.
- [7] Bertsimas, D., Perakis, G. Dynamic Pricing: A Learning Approach. Pp 45-79 in *Mathematical and Computational Models for Congestion Charging*. Book Series: Applied Optimization. D. Hearn and S. Lawphongpanich (eds.), 2006. Springer, June 101, New York.
- [8] Bertsimas, D., Sim, M. The price of Robustness. *SIAM journal of optimization*, 2005, 15:780-804.
- [9] Bertsimas, D., Thiele, A. Robust convex optimization. *Math. Oper. Res.*, 1998, 23:769-805.
- [10] Besbes, O., Zeevi, A. Dynamic Pricing Without Knowing the Demand Function: Risk Bounds and Near-Optimal Algorithms. *Operations Research*, 2009, 57:1407-1420.
- [11] Bitran, G.R., Caldentey, R. An Overview of Pricing Models for Revenue Management. *MSOM*, 2003, 5:203-229.
- [12] Bitran, G.R., Caldentey, Mondschein, S.V. Periodic pricing of seasonal products in retailing. *Management Sci.*, 1997, 43(1):64-79.
- [13] Brafman, R. I., Tenneholtz, M. On the Axiomatization of Qualitative Decision Criteria. *AAAI-97 Proceedings.*, 1997.
- [14] Brémaud, P. Point Processes and Queues, Martingale Dynamics. Springer-Verlag, 1980. New York, NY.

- [15] Borodin, A., El-Yaniv, R. Online Computation and Competitive Analysis. Cambridge University Press, Cambridge, NY.
- [16] Carvalho, A.X., M.L. Puterman. Dynamic pricing and learning over short time horizons. Working Paper, Statistic Department, University of British Columbia, Vancouver, 2004.
- [17] Eren, S., Maglaras, C. Pricing Without Market Information. Working Paper, Columbia University, 2006.
- [18] Farias, V.F., Van Roy, B. Dynamic Pricing with a Prior on Market Response. Working Paper, Stanford University, California, 2007.
- [19] Gallego, G., van Ryzin, G. Optimal Dynamic Pricing of inventories with Stochastic Demand over Finite Horizons. *Management Sci.*, 1994, 40:999-1020.
- [20] Lan, Y., Huina, G., Ball, O.B., Karaesmen, I. Revenue Management with Limited Demand Information. *Management Sci.*, 2009, 54(9):1594-1609.
- [21] Levina, L., Levin, Y., McGil, J., Nediak, M. 2006. Dynamic pricing with online learning of general reservation price distribution. Proceedings of INCOM. Elsevier.
- [22] Lim, A.E.B., Shanthikumar, J.G. Relative entropy, exponential utility, and robust dynamic pricing. *Oper. Res.*, 2006, 55(2):198-214.
- [23] Lin, K.Y. Dynamic pricing with real-time demand learning. *Eur. J. Oper. Res.*, 2006, 174(1):522-538.
- [24] Lobo, M.S., S. Boyd. Pricing and learning with uncertain demand. Working Paper, Duke University, 2003.
- [25] McGill, J.I., van Ryzin, G.J. Revenue Management Without Forecasting or Optimization: An Adaptive Algorithm for Determining Airline Seat Protection Levels. *Management Sci.*, 2000, 46(6):760-775.
- [26] Perakis, G., Roels, G. Robust Controls for Network Revenue Management. *M&SOM*, 2010, 12(1):56-76.
- [27] Seiden, S. S., A Guessing Game and Randomized Online Algorithms. *Proc. 42nd ACM Sympos. Theory of Comput. (STOC2000)*, Portland, OR., 2000, 592-601.
- [28] Smith, S.A., Achabal, D. Clearance Pricing and Inventory Policies for Retail Chains. *Management Sci.*, 1998, 44(3):285-300.
- [29] Talluri, K., van Ryzin, G. The theory and practice of revenue management. Kluwer Academic Press, 2004. Boston, MA.
- [30] Thiele, A. Single-Product Pricing via Robust Optimization. Working paper, Lehigh University, 2006.